



A Practical Introduction to Traffic Flow Theory and On-ramp Flow Control

California Department of Transportation

District 11 – San Diego

A Practical Introduction to Traffic Flow Theory and On-ramp Flow Control

The Staff of D11-Ramp Metering Branch
California Department of Transportation
District 11 – San Diego

Table of Contents

Introduction	4
Traffic Flow Theory	5
Kinematic Waves Model of Traffic Flow	9
Mathematical Solution to Kinematic Waves	12
Practical Traffic Engineering Applications	14
Merging Traffic	15
Dropped Lane or Lane Convergence and Bottlenecks	16
Traffic Management and the Inverse Functions.....	18
Variable Speed Signs.....	21
Conclusion	22

Table of Figures

Figure 1 -	5
Figure 2 -	7
Figure 3 -	8
Figure 4 -	12
Figure 5 -	13

Introduction

In August of 2014, The Transportation Research Board (TRB) Committee on Traffic Flow Theory and Characteristics held a symposium celebrating 50 Years of Traffic Flow Theory in the Midyear Meeting in Portland, Oregon, USA. The predecessor committee focusing on traffic flow theory was organized 50 years ago, to recognize the past accomplishments in the field, reflecting on the present state of our research community and identifying key future directions.

The fact of the matter was that this meeting, like other similar meetings, did not advance the science first proposed by Greenshield in 1930's, and since then a myriad of academics and professionals have expounded on his seminal work and offered as many traffic flow theories and conjectures. These theories, despite their mathematical elegance are mostly inaccurate, as their progenitors had paid no heed to Leibniz's warning that utilizing his mathematical notation always leads to self-consistent but not necessarily accurate results.

In this book we will not attempt to present any of these theories other than what is needed in the minimum for the reader to gain a complete understanding of the traffic flow.

As far as this author is concerned the traffic theory has had three major contributions since Greenshield's original effort. In 1953 Lighthill and Whitman followed by Roberts introduced the LWR model without the required extension into the synchronous traffic movement. In 1964 Newell proposed his model of Fundamental Traffic Diagram without successfully deriving the slope of the downward portion of diagram that corresponds to congested traffic. The unification of LWR and Newell is the hinge pin of understanding the traffic flow and its management as it would be required by the increasing number of vehicles worldwide and the inability of the governments to keep up with the construction of the infrastructure required to support the number of vehicles being used.

Traffic Flow Theory

The most important concept in traffic flow is that of density. The vehicular density is measured in number of vehicles per mile, but in actual engineering practice one does not need to count the number of vehicles in one mile, the distance between two consecutive vehicles suffices to determine density.

$$n = \frac{5280}{L + D}$$

The above equation simply describes the number of vehicle and the distance in front of them that fit inside a mile. We represent that as n and call it vehicular density – 5280 represent the number of feet in one mile, L is the typical vehicular length and D the distance between two consecutive vehicles.

It is important for the reader to realize that in the course of one mile one may encounter several pockets of different vehicular densities, for example two consecutive stop lights may create two pockets of very high density with scant vehicular density in between, the same may be imagined on a roadway where one might encounter vehicles moving in packets of different densities.

Solving for D and dividing by the speed v we may arrive at an equation for the time of impact τ , if the vehicle ahead would come to sudden stop.

$$\tau = \frac{D}{v} = \frac{5280 - nL}{nv}$$

For standard vehicular length of 14.75 ft the Distance versus Density is plotted in Figure 1.

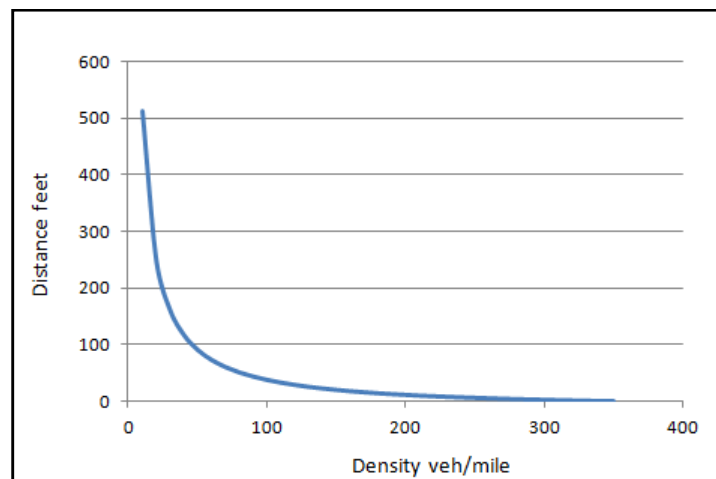


Figure 1 – add text to explain the graphic (if you want)

Conversely, we may also write an equation for speed in miles per hour

$$v = \frac{5280 - nL}{n\tau\bar{c}}$$

Where \bar{c} is $\frac{5280}{3600}$ the conversion factor from ft/sec to MPH.

For ordinary traffic situations, the drivers normally abide by the posted speed limit, therefore we may presume v to be a constant v_f .

We may thus calculate the hourly volume flow of vehicles Q as

$$Q = nv_f$$

It becomes obvious that the time of impact will become increasingly shorter with the increased density until the density reaches a point that the driver chooses to sacrifice the speed in favor of a safe time of impact τ_s , and we may say that at that point we have reached a critical density n_c . Careful measurements of τ_s in various regions have shown that τ_s varies from region to region but it is regionally stable. In southern California τ_s is measured to be 1.75 seconds and in Seoul, Korea it is measured at 2.5 seconds. The speed required to maintain a safe time of impact is then:

$$v = \frac{5280 - nL}{n\tau_s\bar{c}}$$

We may compute n_c from the intersection of the line

$$Q = nv_f$$

with

$$Q = nv = n \frac{5280 - nL}{n\tau_s\bar{c}}$$

Solving for n we get an expression for n_c

$$n_c = \frac{5280}{L + v_f\tau_s\bar{c}}$$

This derivation is universal and applies to vehicular traffic (cars, bicycles, etc.) or even pedestrians and movement of ants. The results of a wonderful graphic experiment can be found at http://tft.ceng.calpoly.edu/tft50/tft50_papers/P08_Seyfried.pdf.

As such n_c becomes a dividing point to bifurcate the traffic flow into two regions of free flow and synchronized flow, a plot of which is presented in Figure 2.

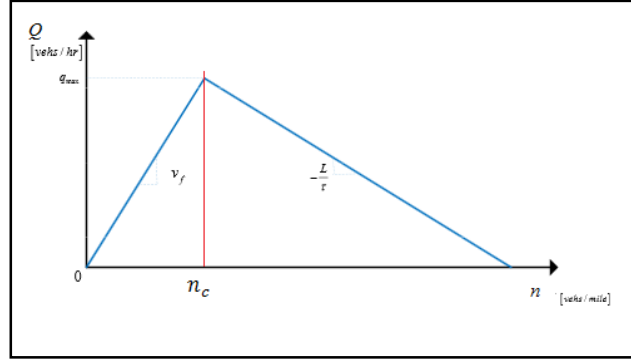


Figure 2 – Fundamental traffic Diagram

The diagram was first proposed by Newell in 1964 and was derived by this author in 2012. What is noteworthy in this diagram is the downward slope of diagram that is computed to be $-\frac{L}{\tau_s}$.

Rewriting the equation for flow when $n > n_c$ as

$$Q = nv = n \frac{5280 - nL}{n\tau_s \bar{c}} = \frac{5280}{\tau_s \bar{c}} - \frac{nL}{\tau_s \bar{c}}$$

We define the first term as total capacity Q_{max} of the freeway

$$Q_{max} = \frac{5280}{\tau_s \bar{c}}$$

And the flow that is not getting through or backing up as

$$q_{backed\ up} = \frac{nL}{\bar{c}\tau_s}$$

Such that the observed flow

$$Q_{manifest} = Q_{max} - q_{backed\ up}$$

And the observed speed will be

$$v_{manifest} = \frac{Q_{manifest}}{n}$$

It is noteworthy that in a situation like NASCAR where τ_s is measured in milliseconds the observed flow will be more than 20,000 vehicles/hour.

Also we need to mention that for $n = n_c$ there is a residual back up flow of $\frac{-n_c L}{\tau_s \bar{c}}$.

A plot of backed up flow and manifest flow is presented below (Figure 3) as a function of density.

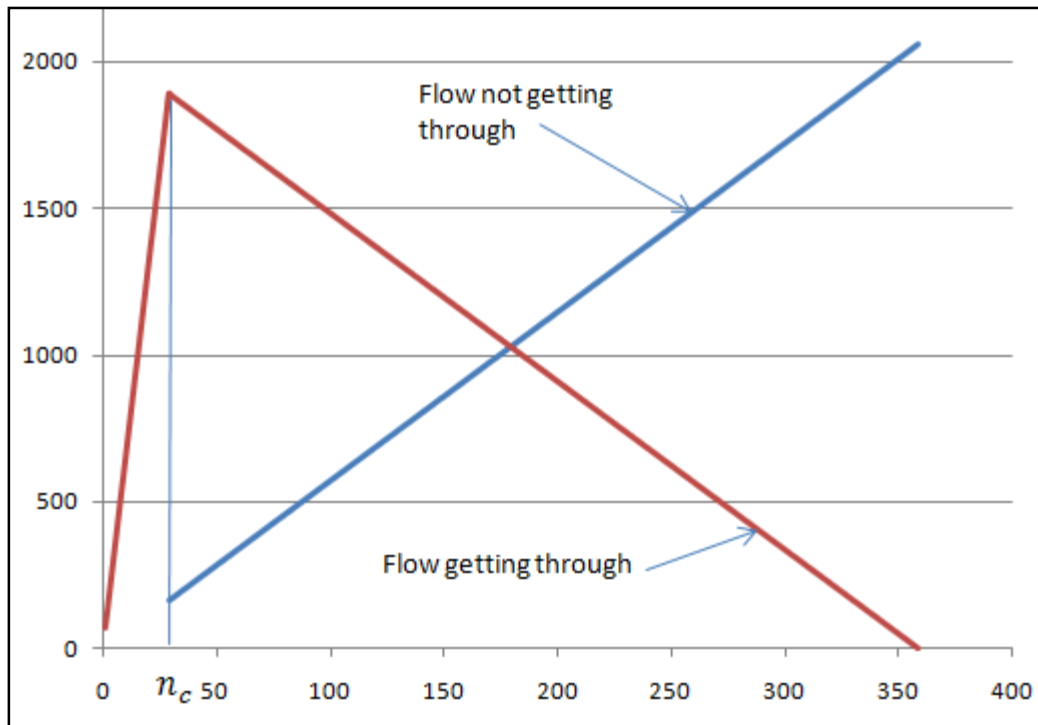


Figure 3 -

The reader may be surprised by the terminology we adopted for which we need to introduce the **Kinematic Waves Theory of Traffic Flow**.

Kinematic Waves Model of Traffic Flow

The kinematic waves theory was seminally proposed by Lighthill and Witham (1955) while studying the tidal backflow into the mouth of river Thames, and later on independently by Roberts in 1956 as a fluid model for traffic flow now known as LW R kinematic waves model. The central postulate of the theory is that there exists a relationship between the flow and the density such that

$$\frac{\partial n}{\partial x} + v \frac{\partial n}{\partial t} = 0$$

The derivation of the above equation is fairly straight forward through the application of first principles of calculus. Here we will repeat the derivation and then extend it to cover both flow regimes of traffic.

The variation of number of vehicles N in the system between two measuring stations A at upstream and B at downstream with time may be written as

$$\frac{dN(t)}{dt} = q_B(B, t) - q_A(A, t)$$

On the other hand we may express N in terms of density

$$N(t) = \int_A^B n(x, t) dx$$

Substituting

$$\frac{dN(t)}{dt} = \frac{d}{dt} \int_A^B n(x, t) dx = q_B(B, t) - q_A(A, t)$$

By Leibniz axiom we may write $q_B(B, t) - q_A(A, t)$ as

$$\int_A^B \frac{\partial}{\partial x} q(x, t) dx$$

Substituting

$$\frac{d}{dt} \int_A^B n(x, t) dx = \int_A^B \frac{\partial}{\partial x} q(x, t) dx$$

Dropping the integrals, we may write the above as

$$\frac{\partial n(x, t)}{\partial t} - \frac{\partial q(x, t)}{\partial x} = 0$$

But $q(x, t) = n(x, t)v(n)$ with v itself having dependence on n , i.e., $v = v(n(x, t))$

$$\frac{\partial n(x, t)}{\partial t} - v(n) \frac{\partial n(x, t)}{\partial x} - n(x, t) \frac{\partial v(n)}{\partial x} = 0$$

Expanding the last term

$$\frac{\partial n(x, t)}{\partial t} - v(n) \frac{\partial n(x, t)}{\partial x} - n(x, t) \frac{\partial v(n)}{\partial n} \frac{\partial n(x, t)}{\partial x} = 0$$

Substituting

$$v(n) = \max \left[v_f, \frac{5280 - nL_{veh}}{n\tau} \right]$$

And

$$\frac{\partial v(n)}{\partial n} = \min \left[0, \frac{5280}{n^2\tau} \right]$$

Applying the two flow regime concepts

$$\frac{\partial v}{\partial n} = 0 \text{ and } v = v_f \forall n < n_c$$

We end up with a wave equation of the form

$$\frac{\partial n(x, t)}{\partial t} - [v_f] \frac{\partial n(x, t)}{\partial x} = 0 \quad \forall n < n_c$$

Which describes a density wave travelling in forward direction with the speed $[v_f]$

And as for the case where

$$v = \frac{5280 - nL_{veh}}{n\tau} \quad \forall n > n_c$$

We may write

$$\frac{\partial n}{\partial t} - \left[\frac{5280 - nL_{veh}}{n\tau} - \frac{5280}{n^2\tau} n \right] \frac{\partial n}{\partial x} = 0$$

Leading to

$$\frac{\partial n(x, t)}{\partial t} + \left[\frac{L_{veh}}{\tau} \right] \frac{\partial n(x, t)}{\partial x} = 0 \quad \forall n > n_c$$

Which describes a wave travelling at speed of $\left[\frac{L_{veh}}{\tau} \right]$ backward, consistent with observations of scientists at the Nagoya University in Japan [<http://iopscience.iop.org/1367-2630/15/10/103034/article>]

The appearance of $\left[\frac{L_{veh}}{\tau} \right]$ unifies the Newell model with LWR fluid model.

Additionally the use of the backed up flow is justified in our previous discussion. The phenomenon is observed daily and has been called the phantom wave by some.

Mathematical Solution to Kinematic Waves

The solution to the Kinematic wave equations introduced in previous chapter is any arbitrary function of the form $n(x, t) = (ax + bvt)$.

However we need to choose our function carefully. Fully knowing that the expected solution is a wave, let us try $a = b = \frac{1}{L+D}$ with a wave length of $L + D$

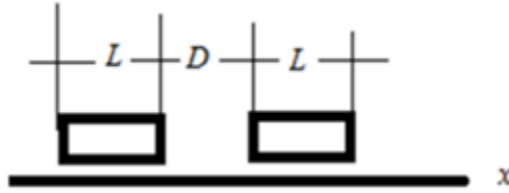


Figure 4 -

Thus we may write

$$n(x, t) = \frac{x}{L + D} + \frac{vt}{L + D}$$

Substituting for D in terms of density from definition of density $n = \frac{5280}{L+D}$, and recognizing the product of $v_f n = \bar{c}Q$ and similarly the product of $\frac{L}{\tau_s} n = \bar{c}q$, we write the solution to the first and the second differential equation as:

$$\frac{nx}{5280} + \frac{Qt}{3600} \quad \forall \quad n < n_c$$

And

$$\frac{nx}{5280} - \frac{qt}{3600} \quad \forall \quad n > n_c$$

Where the first term describes the shape density shape and second term describes transport of it.

The number of vehicles that are accumulated in the backup in the time t_{acc} is then

$$\frac{qt_{acc}}{3600}$$

and the time it will take to disperse the accumulated backed up traffic is then

$$t_{dispersion} = \frac{qt_{acc}}{N 3600} \tau$$

where N is the number of lanes. As expected the time of dispersion is longer than the time of accumulation by the factor of τ .

For simulation purposes, the shape may easily be represented by stringing together (with a minus sign) several boxcar function $\beth(x)$ represented by the Hebrew letter **beth** which is a combination of two Heaviside functions representing a **single vehicle** with its back at location $x = a$ and it's front at location $x = b$,

$$\beth(x) = H(x + a) - H(x - b)$$

Where

$$\beth(x) = \frac{1}{2} [\text{sgn}(x + a) - \text{sgn}(x - b)]$$

Where sgn is the sign function defined as

$$\text{sgn}(x) = \frac{x}{|x|} \quad \forall x \neq 0$$

Then one can make it travel to the right

$$\beth(x) = \frac{1}{2} [\text{sgn}(x + a - vt) - \text{sgn}(x - b - vt)]$$

We have formed a box car function representing 4 vehicles with the density of 30 vehicles per mile in Figure 5.

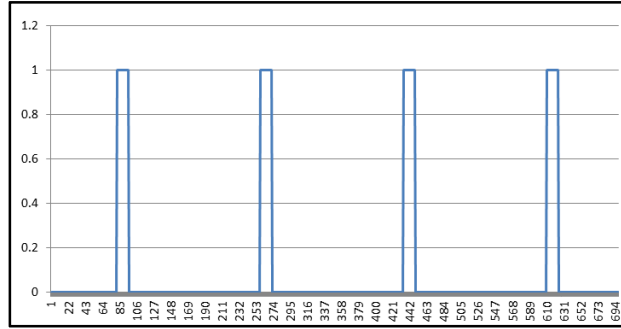


Figure 5 –

The words shape and transport may have come as a surprise to the reader, a brief explanation should clear the matter. Lets imagine a ceramic tiled roof line, if it is not moving it can be described as **wavy**, but not a **wave** since the requirement for a wave is that it will be moving. Now if the roofline can be made to move it will become a moving wavy shape. Similarly we have a string of vehicles with the space in between arranged such that a certain density say 30veh/mile is satisfied, but they are not moving yet. To make them move we need to add the transport term which describes the number of vehicles that get past a point per second. For example if $Q = 360 \text{ veh/hr}$ then a vehicle passes you every 10 seconds and when if $Q = 1800 \text{ veh/hr}$ then a vehicle passes you every 2 seconds. A travelling wave may appear to be stationary, a phenomenon that can be seen by the interference of two waves. The matter may become for clear when the reader attempts to simulate the phenomenon for himself using standard tools such as an Excel sheet.

Practical Traffic Engineering Applications

The practitioners of traffic engineering are more likely interested to solve ordinary day to day traffic engineering problems. Looking at [Figure 5](#) it is easy to imagine how two streams of traffic merge with each other as if each vehicle was a tooth on a zipper. We may at this point debunk all traffic engineering manuals, books and papers prescribing flow additions and affirm that **flows are not additive** and that **only density is additive**, while simultaneously take advantage of the constancy of τ_s and provide a new method to compute the state equation for traffic flow.

Although traffic waves travelling forward and backward are interesting as a phenomenon, we can take advantage of the constancy of impact time and present a new and comprehensive methodology for practitioners of traffic engineering who deal on day to day basis with congestion.

Suppose the traffic undergoes a state change from state $\{A\}$ with density n_a and speed v_a to state $\{B\}$ with density n_b and speed v_b

If $n_b < n_c$ the critical density, then $v_b = v_f$, however if $n_b > n_c$

Then the inter-vehicular distance for state $\{A\}$ may be computed as

$$D_a = \frac{5280 - n_a L}{n_a}$$

Similarly

$$D_b = \frac{5280 - n_b L}{n_b}$$

$$\tau_s = \frac{D_a}{v_a} = \frac{D_b}{v_b}$$

Eliminating τ_s from the above equations, we may write

$$v_b = v_a \frac{D_2}{D_1}$$

And the flow will be

$$Q_b = n_b v_b$$

We may also take advantage of our findings and solve the equation for merging traffic, as well as the added or dropped lanes. See the next section.

Merging Traffic



An on-ramp traffic is merged with the outermost lane. Merging occurs at the speed of the outermost lane with q_r representing on-ramp volume, q_1 and v_1 representing volume and speed on the outermost freeway lane.

The density of the outermost lane can be calculated as

$$n_1 = \frac{q_1}{v_1}$$

The inter vehicular distance in the outermost lane is

$$D_1 = \frac{5280 - n_1 L_{veh}}{n_1}$$

The density of the ramp is

$$n_r = \frac{q_r}{v_1}$$

The total density at the merge is then

$$n_{merg} = n_1 + n_r$$

When $n_{merg} > n_c$ one may compute the inter vehicular distance at merging point

$$D_{merg} = \frac{5280 - n_{merg} L_{veh}}{n_{merg}}$$

The speed after merging

$$v_{merg} = v_1 \frac{D_{merg}}{D_1}$$

The flow after merging

$$q_{merg} = (n_1 + n_r) v_{merg} = n_{merg} v_{merg}$$

Similarly the problem of contraction or expansion in the number of lanes may be addressed.

Dropped Lane or Lane Convergence and Bottlenecks



Let N_u and N_d represent the number of lanes upstream and downstream for a section of the highway. The resulting density after the transition

$$n_d = \frac{N_u}{N_d} n_u$$

Where n_d and n_u are downstream density and upstream density values.

When $n_d > n_c$ one may compute the inter vehicular distance at merging point as

$$D_d = \frac{5280 - n_d L_{veh}}{n_d}$$

And downstream speed as

$$v_d = v_u \frac{D_d}{D_u}$$

Then the downstream flow

$$q_d = n_d v_d$$

The traffic engineering practitioners may be more interested in calculating the time of congestion building. The flow rate at any segment in the system is

$$Q = (q_{upstream} - q_{downstream})$$

The number of vehicles accumulating per unit time (here in seconds) will be

$$N = \frac{Q}{3600}$$

The distance accumulated in unit time is then

$$\frac{dx}{dt} = N(L + D_d)$$

Substituting the corresponding values

$$\frac{dx}{dt} = \frac{(q_{upstream} - q_{downstream})(L_{veh} + D_d)}{3600}$$

By definition

$$(L_{veh} + D_d) = \frac{5260}{n_d}$$

Substituting

$$\frac{dx}{dt} = \frac{(q_{upstream} - q_{downstream})\bar{c}}{n_d}$$
$$\frac{dx}{dt} = \frac{v_{downstream}(q_{upstream} - q_{downstream})\bar{c}}{q_{downstream}}$$

Then the distance x at which the upstream space is depleted in time t

$$x = \frac{1}{N_u} N(L_{veh} + D_d) \times t$$

Replacing $(L_{veh} + D_d)$ and N with their equivalents, the rate of depletion of space will be proportional to the speed of upstream vehicles and is

$$\frac{dx}{dt} = \frac{\bar{c}}{N_u} \frac{(q_{upstream} - q_{downstream})}{n_d} = \frac{v_{downstream}(q_{upstream} - q_{downstream})\bar{c}}{N_u q_{downstream}}$$

The time in seconds for the upstream densities, and speed to equalize with downstream density and speed is

$$t = \frac{3600XN_u q_{downstream}}{v_{downstream}(q_{upstream} - q_{downstream})}$$

Where X, the segment length, is the distance between the upstream and the downstream measurement stations in miles.

Conversely, when $Q = q_{upstream} - q_{downstream}$ is a negative number, it implies a reduction in density as a result of net gain in space.

Traffic Management and the Inverse Functions

The practitioners of traffic engineering may be more interested to discover how they can influence the outcome of traffic flow through management of access to the on-ramps. To accomplish this one needs to predicate on the control of one of the traffic flow variables.

The traffic engineering practitioner may choose to control the density, flow volume, or the queue on the ramp. In the following discussions we have provided a prescription for control of these variables via the inverse functions.

$$n_{merge} = n_{ramp} + n_{upstream}$$

Substituting for their values

$$n_{merge} = \frac{q_{ramp}}{v_{upstream}} + \frac{Q_{upstream}}{v_{upstream}}$$

The merge density of the traffic on the main lanes after a flow from the ramp to an N lane facility is

$$n_{merge} = \frac{q_{ramp}}{Nv_{upstream}} + \frac{Q_{upstream}}{v_{upstream}}$$

If controlling the density is a goal we utilize the above equation by asserting $n_{merge} = n_{goal}$

$$n_{goal} = \frac{q_{ramp}}{Nv_{upstream}} + \frac{Q_{upstream}}{v_{upstream}}$$

$$q_{ramp} = N(n_{goal}v_{upstream} - Q_{upstream})$$

Where N is the number of the main lanes. A negative value describes a flow rate that needs to be redirected out of freeway to accomplish the goal. As such predicating on density is not a good choice for traffic control since we have no control over the number of vehicles that need to be redirected out of the system.

A better choice to maintain a sustainable flow gradient to optimize travel time is the control of flow on the main lanes for a sustainable flow gradient. The derivation is tedious but straight forward.

The speed of the merge may be described by the equation of merge

$$v_{merge} = v_{upstream} \frac{D_{merge}}{D_{upstream}}$$

We may rewrite the above in terms of respective density

$$v_{merge} = v_{upstream} \frac{5280 - \left[\frac{q_{ramp}}{N v_{upstream}} + \frac{Q_{upstream}}{v_{upstream}} \right] L}{\frac{q_{ramp}}{N v_{upstream}} + \frac{Q_{upstream}}{v_{upstream}}} \frac{5280 - \left[\frac{Q_{upstream}}{v_{upstream}} \right] L}{\frac{Q_{upstream}}{v_{upstream}}}$$

And multiplying by

$$n_{merge} = \frac{q_{ramp}}{v_{upstream}} + \frac{Q_{upstream}}{v_{upstream}}$$

and a bit of algebraic manipulation we arrive at

$$Q_{merge} = Q_{upstream} - \frac{n_{upstream} L q_{ramp}}{N(5280 - n_{upstream} L)}$$

Where N is the number of main lanes.

Rewriting the above

$$q_{ramp} = N \frac{(5280 - n_{upstream} L)}{n_{upstream} L} (Q_{upstream} - Q_{merge})$$

$$q_{ramp} = N \frac{(5280 - n_{upstream} L)}{n_{upstream} L} (\Delta Q)$$

And since the density data is not readily available one may substitute its equivalent in terms of measured loop occupancy. In California where 10 ft. detection loops are prevalent the conversion factor is

$$n = 233.333 \text{ } 0cc$$

The ramp discharge rate then is

$$q_{ramp} = N \frac{(5280 - 213.330cc_{upstream} L)}{213.330cc_{upstream} L} (\Delta Q)$$

The practitioner may also want to control the queue length on the ramp.

The queue on the on-ramp has to be controlled for efficient operation of the corridor and to minimize the impact of the ramp backup on the local streets. This is achieved by modulating the queue build-up rate in response to the rate of arrival Q_{in} .

$$dq = Q_{out} - Q_{in}$$

The number of vehicles on the ramp at any time t

$$dqt = (Q_{out} - Q_{in})t + \delta_x$$

Where δ_x is the number of vehicles on the ramp before the start of measurement.

Total vehicle length on the ramp

$$L_{veh}(Q_{out} - Q_{in})t + \delta_x L_{veh}$$

Percentage R_{occ} of ramp capacity occupied

$$\frac{L_{veh}(Q_{out} - Q_{in})t + \delta_x L_{veh}}{L_{ramp}} = R_{occ}$$

Where L_{ramp} is the total length of all lanes on the ramp.

Equating $\delta_x = 0$ and solving for Q_{out} provides the prescription for controlling the percentage of the ramp capacity occupied.

$$Q_{out} = Q_{in} - \frac{R_{occ} L_{ramp}}{L_{veh} t} = q_{ramp}$$

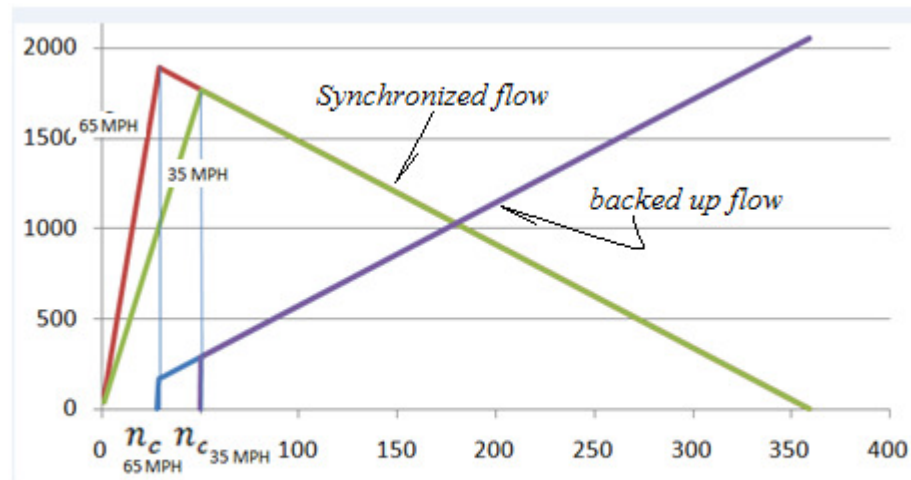
Equating Q_{out} to q_{ramp} we have the algorithm to sustain a flow gradient with the percentage of the occupancy of the ramp capacity as a controlling factor and its impact at the merge point. Some ramps may allow storage capacity to exceed physical capacity, in case of separate lanes for ramp entrance; others may be restricted to 60%, for example. This parameter may be governed by policies and ramp configuration.

Variable Speed Policy

As of late a few experiments (Georgia DOT and EC) has been conducted in support of variable speed limit during the congestion hours. California is also considering the variable speed policy on certain corridors.

A quick look at the graph below shows that all n_c regardless of the posted speed limit v_f lie on the manifest flow (Synchronized flow) line described by

$$Q = \frac{5280}{\tau_s \bar{c}} - \frac{nL}{\bar{c} \tau_s}$$



As such the posted speed plays no role in the behavior of synchronized traffic, one may verify that when the traffic condition is such that one can only drive at say 10 miles per hour, it would not matter whether the sign on the side of the road for speed limit is fixed at 65 MPH or variable at 35 MPH

Conclusion

In this paper we have derived all equations needed for complete understanding of traffic flow. We have offered a sensible solution to the equation of kinematic waves consistent with both the traffic flow diagram and field observation. At the same time we have presented a practical method to compute traffic variables for everyday problems that a traffic engineer may encounter. We have also provided a comprehensive method to control the flow from the ramp to achieve any desired outcome. We have offered a method to estimate the time it takes for a traffic jam to clear. Finally, we have offered a universal approach to balance ramp capacity with its daily use and computed outflow. As a result we have debunked the commonly accepted practice of adding flow rates as is prescribed in current traffic manuals , books and papers, and we have shown that the variable speed limit is not of consequence in controlling the freeway flow.

We expect that the application of these innovative approaches would help with the manifest gridlock that some urban areas are experiencing. We hope that readers will use the knowledge to simulate their local area traffic situation and re-study the effects of applying the appropriate ramp rates without the need for calibrating very expensive and time consuming traffic simulation software.